

Market Research on Caffeinated Products

A start-up company in Boulder has an idea for a new energy product: caffeinated chocolate. Many details about the exact product concept are yet to be decided, however, the company wanted to get some consumer input before it moved ahead with developing the new product. To do that, they created concept statements for seven related ideas, listed below.

The exact statements of the concepts are included in the data workbook for this exercise. (See the tab labeled concepts.) The concepts numbered 1 and 3 (chocolate bar and chocolates) are closest to what the company has in mind for the new product. Concept number 2 is a similar idea, while concepts 4-6 are quite different categories from their focal idea. Concept 7 (hamburger meat) served as a benchmark: it was designed as an intentionally unattractive idea, and it was used to see if people are really paying attention to the survey or just answering randomly.

1. Caffeinated chocolate bar.
2. Caffeinated cookies.
3. Caffeinated chocolates.
4. Caffeinated potato chips.
5. Caffeinated water.
6. Caffeinated cereal/granola.
7. Caffeinated hamburger meat.

The survey first asked some background questions (gender, age, employment status, and usage of energy products) and then asked four questions for each of the seven concepts. The four questions for each concept were 1) purchase intent, 2) uniqueness of the idea, 3) projected usage frequency, and 4) substitution behavior. Each survey respondent saw the concepts in a random order. The actual survey is included with this exercise.

The data from the survey is in the data file for this exercise, on the tab labeled survey-results. There were 150 responses.

Part 1: Purchase Intent

Each respondent was asked a “purchase intent” question for each concept. The wording of the question is, “Based on the description, how likely would you be to purchase this product?” This is a standard question in market research for consumer products.

In the data file, the responses to the purchase intent questions are in columns labeled INTENT(1), INTENT(2), up to INTENT(7), and the numbering of concepts 1-7 is as shown above.

In those columns,

- 1 means the respondent answered Definitely would not,
- 2 is Probably would not,
- 3 is Might or might not,
- 4 is Probably would, and
- 5 is Definitely would.

1) Average Purchase Intent

Using all the people who answered the question for each concept, find the average response for each of the seven concepts. Also find how many people answered the purchase intent question for each concept.

#	Concept Name	Average Purchase Intent	Number of Responses
1	Caffeinated chocolate bar.	2.49	140
2	Caffeinated cookies.		
3	Caffeinated chocolates.		140
4	Caffeinated potato chips.		
5	Caffeinated water.	2.04	139
6	Caffeinated cereal/granola.	1.72	
7	Caffeinated hamburger meat.		142

- a. Which concept has the highest average score for purchase intent?
- b. Which concept has the lowest average score for purchase intent?

2) Distribution of Responses

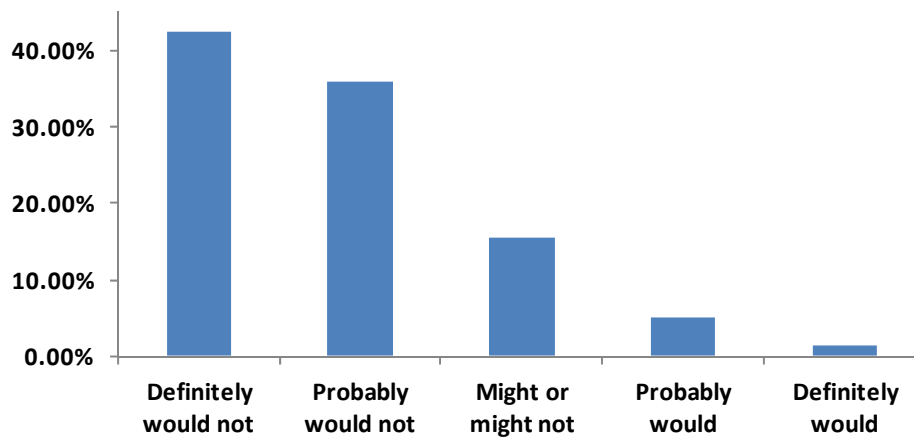
Find the distribution of responses for the concepts below: what percentage of the responses for each product are 1s, 2s, etc.? Give your answers as percentages with two decimal places.

Distribution of Responses for Purchase Intent						
#	Concept Name	% 1s	%2s	%3s	%4s	%5s
2	Caffeinated cookies.	42.25%	35.92%	15.49%		
3	Caffeinated chocolates.		29.29%	27.14%		

3) Graph of Distribution of Responses

Below is a column graph showing the full distribution for caffeinated cookies (concept 2). Create a similar graph in Excel, showing the distribution for caffeinated chocolates (concept 3).

Relative Frequency of Purchase Intent Responses for the Caffeinated Cookies Concept, n = 142



4) Top-Two Boxes

Some companies look at the “top-two boxes” score for concepts. The top-two boxes is simply the percentage of respondents who answered Probably plus the percentage of respondents who answered Definitely. For example, if 6% of respondents answered Definitely (a response of 5 in the data) for one of the concepts and 20% of respondents answered Probably (a response of 4 in the data) for that concept, then the top-two boxes score for that concept is 26%.

a. What is the top-two boxes score for Caffeinated chocolate bar (concept 1)?

b. Which of the seven concepts has the highest top-two boxes score?

Purchase intent responses can be coded as probabilities. One way of translating the response phrases to probabilities is as follows:

- Definitely would not = 0% chance,
- Probably would not = 25% chance,
- Might or might not = 50% chance,
- Probably would = 75% chance, and
- Definitely would = 100% chance.

5) Average Purchase Probabilities

Using these values, what is the average probability of purchase for each of the seven concepts?
(Use all the responses, not just the top-two boxes.)

Average Purchase Probabilities, Using 0%, 25%, 50%, 75%, 100%		
#	Concept Name	Average Purchase Probability
1	Caffeinated chocolate bar.	37.14%
2	Caffeinated cookies.	21.83%
3	Caffeinated chocolates.	
4	Caffeinated potato chips.	
5	Caffeinated water.	
6	Caffeinated cereal/granola.	18.04%
7	Caffeinated hamburger meat.	

- a. Which concept has the highest average probability of purchase?

- b. Which concept has the lowest average probability of purchase?

- c. Which of the following statements is true in general, not just in this data set?

Statement 1: The ranking of concepts by average purchase probability using probabilities of 0%, 25%, 50%, 75%, and 100% is exactly equivalent to the ranking of concepts using average survey responses (where the survey responses are 1, 2, 3, 4, 5).

Statement 2: The ranking of concepts by average purchase probability using probabilities of 0%, 25%, 50%, 75%, and 100% is not exactly equivalent to the ranking of concepts using average survey responses (where the survey responses are 1, 2, 3, 4, 5), but the two methods should often produce similar rankings.

Statement 3: The ranking of concepts by average purchase probability using probabilities of 0%, 25%, 50%, 75%, and 100% is not exactly equivalent to the ranking of concepts using average survey responses (where the survey responses are 1, 2, 3, 4, 5), and the two methods will usually produce dissimilar rankings.

People often overstate their intention to purchase. Instead of 0%, 25%, 50%, 75%, and 100%, a more realistic translation from response phrases to probabilities is as follows:

- Definitely would not = 0% chance,
- Probably would not = 2% chance,
- Might or might not = 10% chance,
- Probably would = 25% chance, and
- Definitely would = 50% chance.

6) Using these values, what is the average probability of purchase for each of the seven concepts?

Average Purchase Probabilities, Using 0%, 2%, 10%, 25%, 50%		
#	Concept Name	Average Purchase Probability
1	Caffeinated chocolate bar.	9.65%
2	Caffeinated cookies.	4.20%
3	Caffeinated chocolates.	
4	Caffeinated potato chips.	
5	Caffeinated water.	
6	Caffeinated cereal/granola.	
7	Caffeinated hamburger meat.	0.86%

- a. Which concept has the highest average probability of purchase?
- b. Which concept has the lowest average probability of purchase?

- c. Which of the following statements is true in general, not just in this data set?

Statement 1: The ranking of concepts by average purchase probability using probabilities of 0%, 2%, 10%, 25%, and 50% is exactly equivalent to the ranking of concepts using average survey responses (where the survey responses are 1, 2, 3, 4, 5).

Statement 2: The ranking of concepts by average purchase probability using probabilities of 0%, 2%, 10%, 25%, and 50% is not exactly equivalent to the ranking of concepts using average survey responses (where the survey responses are 1, 2, 3, 4, 5), but the two methods should often produce similar rankings.

Statement 3: The ranking of concepts by average purchase probability using probabilities of 0%, 2%, 10%, 25%, and 50% is not exactly equivalent to the ranking of concepts using average survey responses (where the survey responses are 1, 2, 3, 4, 5), and the two methods will usually produce dissimilar rankings.

7) Comparing Top-Two Boxes Percentages and Average Purchase Probabilities

TRUE or FALSE: if a concept has the largest top-two boxes score, it will also have the highest average purchase probability. Explain your answer.

Part 2: Frequency

Each respondent was asked a “frequency” question for each concept. The wording of the question is, “If you tried this product and it met your expectations for taste and effectiveness, how often could you see yourself using it?” This is a standard question in market research for non-durable consumer products.

In the data file, the responses to the purchase intent questions are in columns labeled FREQ(1), FREQ(2), up to FREQ(7), and the numbering of concepts 1-7 is as shown at the beginning of the module (and is also included in the tab labeled concepts in the data file included with this exercise).

In those columns,

- 1 means the respondent answered Once a month or less often,
- 2 is More than once a month but less than once a week
- 3 is 1 to 3 times a week,
- 4 is More than 3 times a week, and
- 5 is Don’t know.

1) Distribution of Responses

Find the distribution of responses for the concepts below: what percentage of the responses for each product are 1s, 2s, etc.? Give your answers as percentages with two decimal places.

Distribution of Responses for Frequency						
#	Concept Name	% 1s	%2s	%3s	%4s	%5s
2	Caffeinated cookies.	61.76%		2.94%		23.53%
3	Caffeinated chocolates.	57.46%				11.19%

We can use the frequency responses to estimate annual usage behavior. Similar to our translation of intent responses to probabilities, we can translate the frequency responses to annual usage numbers. Of course, these are just estimates, but they give a reasonable approximation. Use the following values (which can be found on the look-ups tab in the Excel workbook):

- Once a month or less often = 5 times a year
- More than once a month but less than once a week = 15 times a year
- 1 to 3 times a week = 60 times a year
- More than 3 times a week = 100 times a year
- Don’t know = 0 times a year.

Notice that these values are conservative translations, once again assuming that people overstate their intentions.

2) Estimated Annual Usage

Find the estimated annual usage for each concept using those values:

Estimated Annual Usage, Using 5, 15, 60, 100, 0		
#	Concept Name	Estimated Annual Usage
1	Caffeinated chocolate bar.	11.30
2	Caffeinated cookies.	7.87
3	Caffeinated chocolates.	
4	Caffeinated potato chips.	
5	Caffeinated water.	
6	Caffeinated cereal/granola.	8.72
7	Caffeinated hamburger meat.	

3) Averages of Responses

For the purchase intent data, we looked at the average value of the responses. Mathematically, we can also take averages of the responses to the frequency question. Is that a useful thing to do, yes or no?

If yes: what is the interpretation of the average frequency response?

If no: why not?

4) Normalizing: Excluding the Don't Knows

Above, we treated Don't Knows as "won't buys." That is, we estimated 0 annual usage for people who responded Don't Know to the frequency question. That's a somewhat extreme assumption.

- a. Instead, we could just ignore or exclude the Don't Know responses. To do that, we'd look at the percent of each response out of the total number of responses 1-4. Find those percentages and fill in the chart below.

Distribution of Responses for Frequency, Excluding Don't Know (5)					
#	Concept Name	% 1s out of all 1-4 responses	% 2s out of all 1-4 responses	% 3s out of all 1-4 responses	% 4s out of all 1-4 responses
2	Caffeinated cookies.		13.46%	3.85%	
6	Caffeinated cereal/granola.		16.19%	4.76%	

- b. Using the same estimates of frequency for responses 1-4 from above (e.g., a response of 1, which is Once a month or less often, is 5 times a year), find the following:

Estimated Annual Usage, Excluding Don't Know (5)		
#	Concept Name	Estimated Annual Usage
1	Caffeinated chocolate bar.	13.38
2	Caffeinated cookies.	10.29
3	Caffeinated chocolates.	
4	Caffeinated potato chips.	
5	Caffeinated water.	
6	Caffeinated cereal/granola.	11.05
7	Caffeinated hamburger meat.	

- c. The estimated annual usage for each concept is higher when we exclude the Don't Know responses. Why?

Part 3: By Gender and Age

The column labeled GEN has responses to a question about gender. The responses are coded as follows:

- 1 is Male
- 2 is Female

The column labeled AGE has responses to a question about age. The responses are coded as follows:

- 1 means the respondent answered under 18 years old
- 2 is 18-25 years old
- 3 is 26-35 years old
- 4 is 36-45 years old
- 5 is 46-55 years old
- 6 is 56-65 years old
- 7 is 66-75 years old
- 8 is over 75 years old

1) Find the average purchase intent response by male respondents:

Purchase Intent by Male Respondents			
#	Concept Name	Average Purchase Intent	Number of Responses
1	Caffeinated chocolate bar.	2.56	39
2	Caffeinated cookies.		41
3	Caffeinated chocolates.		39
4	Caffeinated potato chips.	1.51	
5	Caffeinated water.	2.13	
6	Caffeinated cereal/granola.	1.56	
7	Caffeinated hamburger meat.		40

2) For which of the seven concepts do males have higher average purchase intent than females?

3) Find the average purchase intent response by respondents 45 and under:

Purchase Intent by Respondents 45 and Under			
#	Concept Name	Average Purchase Intent	Number of Responses
1	Caffeinated chocolate bar.	2.58	97
2	Caffeinated cookies.	1.94	99
3	Caffeinated chocolates.	2.66	
4	Caffeinated potato chips.		
5	Caffeinated water.		
6	Caffeinated cereal/granola.		
7	Caffeinated hamburger meat.	1.22	98

4) For which of the seven concepts do respondents 45 and under have higher average purchase intent than respondents over 45?

Part 4: Correlations with Current Consumption of Energy Bars

As part of the background questions, each respondent was asked about their current usage of energy products. In the column labeled **CUR-USE_5** are the responses to the question, “Rate your consumption for any brand of energy bars (examples: PowerBar, LunaBar, Clif Bar).”

The responses were coded as follows.

- 1 means the respondent answered I've never tried it,
- 2 is I've tried it but consume it less than once a month,
- 3 is I consume it more than once a month but less than once a week,
- 4 is I consume it 1 to 3 times a week, and
- 5 is I consume it more than 3 times a week.

We will look at the data related to the following question: ***Do people who said they consume more energy bars tend to have a higher stated purchase intention for the concepts?***

To explore the answer to that question, compute the correlation coefficient between the responses to CUR-USE_5 and each of the purchase intent questions. The correlation coefficient is a value that gives the strength of the linear relationship between two quantities. Correlation is a real number between -1 and 1. A relatively large positive number (like .8) would mean that there is a strong positive linear relationship between energy consumption and purchase intent—the more energy bars consumed, the higher the purchase intent.

- 1) Find the correlation between the current energy bar consumption responses (reported in CUR-USE_5) and the purchase intent responses for each concept (reported in INTENT(1) through INTENT(7)):

Correlations of Energy Bar Consumption and Purchase Intent		
#	Concept Name	Correlation
1	Caffeinated chocolate bar.	-0.10
2	Caffeinated cookies.	
3	Caffeinated chocolates.	-0.04
4	Caffeinated potato chips.	
5	Caffeinated water.	
6	Caffeinated cereal/granola.	
7	Caffeinated hamburger meat.	-0.08

- 2) Based on the correlations you calculated, would higher energy bar consumption predict higher purchase rates?

- 3) Create a scatter plot showing the relationship between the response to the question about consumption of energy bars (the CUR-USE_5 column) and the response to the purchase intent for the first concept, the caffeinated chocolate bar (the INTENT (1) column). Can you see the negative relationship in the scatter (corresponding to the correlation of -0.10)? Why or why not?

Part 5: Excluding Low Quality Responses

The “caffeinated hamburger meat” concept wasn’t a real concept—it was just a check to see if people were paying attention to the survey. Although it is possible that someone would have a high purchase intent for that concept, the concept was included to help us screen out people who were not taking their responses seriously. When people took the survey, the response for Definitely was the closest to the “advance to the next screen” button, so people who were just trying to speed through the survey would be more likely to give that response.

We also want to screen out people who didn’t answer all the questions.

- 1) Find the average purchase intent responses excluding “low quality” responses. We define a low quality responses as any row in the data set that meets at least one of these criteria:
 - a response of 3, 4, or 5 for purchase intent on concept 7, caffeinated hamburger meat
 - failure to answer all twenty-one INTENT, UNIQ, and FREQ questions (see the column headers in Row 1 of the data).

There are many ways to do the exclusions. Try to avoid using a manual process like deleting each of the low quality rows. (Imagine if the data set had 150,000 rows instead of 150: you wouldn’t want to go one by one through all of them.)

Purchase Intent, Excluding “Low Quality” Responses			
#	Concept Name	Average Purchase Intent	Number of Responses
1	Caffeinated chocolate bar.	2.50	125
2	Caffeinated cookies.		
3	Caffeinated chocolates.		
4	Caffeinated potato chips.		
5	Caffeinated water.	2.01	
6	Caffeinated cereal/granola.	1.74	
7	Caffeinated hamburger meat.		